CS 285 HW 4

**Q1**

**Chart, line chart

Description automatically generated**

Caption: MPE for q1\_cheetah\_n500\_arch1x32

Chart, diagram

Description automatically generated

Caption: MPE for q1\_cheetah\_n5\_arch2x250

Chart, line chart

Description automatically generated

Caption: MPE for q1\_cheetah\_n500\_arch2x250

**Comment:** The last setup with 2-layer, 250-size MLPs as the dynamic models perform the best (although it is close to the first setup), since the mean MPE loss is slightly smaller than that of the first setup. The superior performance may be attributed to the greater expressiveness of the larger dynamic model, and the fact that there are a much greater number of agents trained per iteration as compared to the second setup.

**Q2**

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

Caption: The training average return (about -164) and the evaluation average return (about -43) for q2. They are all within or better than the expectation.

**Q3**

Chart, line chart

Description automatically generated

Caption: The evaluation return of the obstacle env for q3. The final evaluation average return is around -26, as expected.

Chart, line chart

Description automatically generated

Caption: The evaluation return of the reach env for q3. The final evaluation average return is about -250, as expected.

Chart, line chart

Description automatically generated

Caption: The evaluation return of the cheetah env for q3. The final evaluation average return is about 358, as expected.

**Q4**

**Chart, line chart

Description automatically generated**

Caption: I observe from the above plot that a relatively short horizon for planning actions gives the best average evaluation return, although its initial evaluation return is poor.

**Chart, line chart

Description automatically generated**

Caption: I observe from the above plot that a larger number of random action candidate sequences gives a better performance. As can be seen, the yellow curve corresponding to 1000 number of sequence considered per action works better.

**Chart, line chart

Description automatically generated**

Caption: I observe from the above plot that a larger ensemble size for the dynamics model gives a better performance. As can be seen, the green curve corresponding to an ensemble size of 5 works the best among the three.